

**МІНІСТЕРСТВО ОСВІТИ І НАУКИ, МОЛОДІ ТА СПОРТУ УКРАЇНИ
НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ»**

ФІЗИКО- ТЕХНІЧНИЙ ІНСТИТУТ

Кафедра інформаційної безпеки

КОМП'ЮТЕРНИЙ ПРАКТИКУМ №1

з дисципліни

Криптографія

З теми: « Експериментальна оцінка ентропії на символ джерела відкритого тексту »

Перевірила:

Селюх П.В

Виконав студент групи ФБ-92

Андрієвич Дмитро Юрійович

Мета роботи: Засвоєння понять ентропії на символ джерела та його надлишковості, вивчення та порівняння різних моделей джерела відкритого тексту для наближеного визначення ентропії, набуття практичних навичок щодо оцінки ентропії на символ джерела.

Завдання:

0. Уважно прочитати методичні вказівки до виконання комп'ютерного практикуму.
1. Написати програми для підрахунку частот букв і частот біграм в тексті, а також підрахунку H_1 та H_2 за безпосереднім означенням. Підрахувати частоти букв та біграм, а також значення H_1 та H_2 на довільно обраному тексті російською мовою достатньої довжини (щонайменше 1Мб), де імовірності замінити відповідними частотами. Також одержати значення H_1 та H_2 на тому ж тексті, в якому вилучено всі пробіли.
2. За допомогою програми CoolPinkProgram оцінити значення $H(10)$, $H(20)$, $H(30)$.
3. Використовуючи отримані значення ентропії, оцінити надлишковість російської мови в різних моделях джерела

1.

Код програми знаходиться у файлах **main.py** та **text_processing.py**.

Результати:

H_1 with whitespaces: 4.369094903350895

H_2 with whitespaces: 2.7865215549670412

H_1 without whitespaces: 4.388593990853912

H_2 without whitespaces: 3.1294773148931543

Character frequency without whitespaces

Character	Frequency
о	0,109356
и	0,092017
е	0,081075
а	0,080222
т	0,074005
н	0,065514

с	0,055553
л	0,047091
р	0,046152
к	0,040951
в	0,036505
п	0,030713
з	0,029632
м	0,025388
д	0,023296
я	0,021679
г	0,020216
ь	0,016737
й	0,014079
ю	0,013831
б	0,012615
у	0,01212
ж	0,010509
ч	0,009952
ы	0,009818
х	0,006947
ш	0,004324
щ	0,004146
э	0,002773
ф	0,001397
ц	0,001388

Character frequency with whitespaces

Character	Frequency
space	0,133843
о	0,094719
и	0,079701
е	0,070224
а	0,069485

т	0,0641
н	0,056745
с	0,048117
л	0,040789
р	0,039975
к	0,03547
в	0,031619
п	0,026602
з	0,025666
м	0,02199
д	0,020178
я	0,018777
г	0,017511
ь	0,014496
й	0,012195
ю	0,01198
б	0,010926
у	0,010498
ж	0,009102
ч	0,00862
ы	0,008504
х	0,006018
ш	0,003745
щ	0,003591
э	0,002401
ф	0,00121
ц	0,001202

Bigram frequency with whitespaces (**bigram_frequency_with_whitespaces.xlsx**)

	а	б	в	г	д	е	ж	з	и	й	к	л	м	н	о	п	р	с	т	у	ф	х	ц	ч	ш	щ	ы	ь	э	ю	я			
а	0	0	0,004867	0	0,002389	0,004953	0,006681	0,004909	0	0	0,007473	0,007267	0,002421	0,003622	0	0	0,001135	0,008047	0,004887	0	0	0	0	0	0	0	0	0	0	0,003576	0			
б	0	0	0	0	0	0,002535	0	0	0,001135	0	0	0	0	0	0,002452	0	0	0,001164	0	0	0	0	0	0	0	0	0	0,003641	0	0	0	0		
в	0,007251	0	0	0	0	0,004949	0	0	0,001198	0	0,001247	0	0,001198	0,010936	0	0,001258	0,001202	0	0	0	0	0	0	0,001239	0	0	0,001142	0	0	0	0	0		
г	0,004909	0	0	0	0,001317	0	0	0	0	0	0	0	0	0	0,009998	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
д	0,007252	0	0	0	0	0,006738	0	0	0	0	0,001251	0	0	0	0	0,001216	0	0,001247	0	0	0	0	0	0,001265	0	0	0	0	0	0	0	0		
е	0	0	0,001258	0,00375	0,006094	0,001152	0	0,001176	0	0,006036	0,004837	0,006073	0,002452	0,007989	0	0,001195	0,003611	0,003627	0,007386	0	0	0,001228	0	0	0	0	0	0	0	0	0	0		
ж	0,00121	0	0	0	0,005512	0	0	0	0,001211	0	0	0	0	0,001169	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
з	0,00367	0,002535	0,001251	0	0	0	0	0,00121	0,001256	0,002386	0	0	0,001135	0	0,003625	0,002491	0	0	0	0,001187	0	0	0	0	0	0	0	0,001221	0,001227	0	0	0		
и	0	0	0,002534	0,001172	0	0,006125	0	0,007441	0	0,00244	0,008573	0,004737	0,008652	0,00601	0	0	0,001176	0,001187	0,001198	0	0	0,003615	0,001202	0,00121	0,001284	0	0	0	0,001236	0,004126	0	0		
й	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0,001198	0	0	0	0	0	0	0	0,001222	0	0	0	0	0	0		
к	0,008542	0	0	0	0	0	0	0	0,006084	0	0	0	0	0	0,001149	0,011274	0	0,002431	0	0,002297	0,001164	0	0	0	0	0	0	0	0	0	0	0		
л	0,004302	0	0	0	0	0,001247	0	0	0,009821	0	0	0	0	0	0,006214	0	0	0	0	0	0	0	0	0	0	0	0	0	0,007208	0	0,002372	0,001175		
м	0,002471	0	0	0	0	0,003624	0	0	0,006153	0	0	0	0	0	0,001187	0,002444	0	0	0	0,001208	0	0	0	0	0	0	0	0	0	0	0	0,001215		
н	0,006825	0	0	0	0	0,009715	0	0,020136	0	0	0	0	0	0,001271	0,011596	0	0	0	0	0,00347	0	0	0	0	0	0	0	0,001261	0	0	0,001227	0,001244		
о	0	0,003579	0,00376	0,008883	0,001265	0,006197	0	0,004758	0	0,003719	0	0,004846	0,002409	0,001244	0	0	0,00498	0,012195	0,008608	0	0	0	0	0,001216	0	0	0	0	0	0	0,001192	0,001164		
п	0	0	0	0	0	0,001208	0	0	0	0	0	0	0	0	0,00966	0	0,015734	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
р	0,01329	0	0	0,001202	0	0,00964	0	0	0	0,007321	0	0	0	0	0	0,006148	0	0	0,001135	0	0	0	0	0	0	0	0	0	0	0,001239	0	0	0	
с	0,001158	0	0	0	0	0	0	0	0,001215	0	0,002401	0,007929	0	0,001227	0,001228	0,002477	0	0	0,018168	0	0	0,001175	0	0,001282	0	0	0	0	0	0,001202	0,00611	0		
т	0,004949	0	0,003564	0	0	0,006007	0	0,004745	0	0,002403	0,001282	0	0,003713	0,01551	0	0,003604	0	0	0	0	0	0	0	0	0	0	0	0	0,004778	0	0	0,001261		
у	0	0	0	0	0,001164	0	0	0	0	0	0	0	0,002339	0	0	0,002317	0	0	0,001149	0	0	0	0	0	0	0	0	0	0	0	0	0		
ф	0	0	0	0	0	0	0	0	0,00121	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
х	0	0	0,001175	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
ц	0,001202	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
ч	0,001282	0	0	0	0	0,003717	0	0	0,001265	0	0	0	0	0	0	0	0	0	0,002355	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
ш	0	0	0	0	0	0	0	0	0,002461	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0,001284	0	0	0	0		
щ	0	0	0	0	0	0,001215	0	0	0,002377	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
ы	0	0	0,001239	0	0	0	0	0	0	0	0	0	0,001256	0,001261	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
ь	0	0	0	0	0	0	0	0,002446	0	0	0,001176	0	0	0,001198	0	0	0	0,004794	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
э	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0,002401	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
ю	0	0,001135	0	0	0,001238	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0,002374	0	0	0	0	0	0	0,002377	0	0	0	0	0	0	0
я	0	0	0	0,001261	0	0	0	0	0	0	0	0	0,001244	0,001169	0	0	0	0,001227	0,003624	0	0	0	0	0	0	0	0	0,001215	0	0	0	0,001175	0	

Bigram frequency without whitespaces (bigram_frequency.xlsx)

	а	б	в	г	д	е	ж	з	и	й	к	л	м	н	о	п	р	с	т	у	ф	х	ц	ч	ш	щ	ы	ь	э	ю	я			
а	0	0	0,00562	0	0,002758	0,005719	0,007713	0,005667	0	0	0,008628	0,00839	0,002796	0,004182	0	0	0,00131	0,009291	0,005643	0	0	0	0	0	0	0	0	0	0	0,004129	0			
б	0	0	0	0	0	0,002927	0	0	0,00131	0	0	0	0	0	0,002831	0	0	0,001344	0	0	0	0	0	0	0	0	0	0,004203	0	0	0	0		
в	0,008371	0	0	0	0	0,005714	0	0	0,001383	0	0	0,001439	0	0,001383	0,012626	0	0,001452	0,001388	0	0	0	0	0	0	0,001431	0	0,001319	0	0	0	0	0		
г	0,005667	0	0	0	0,001521	0	0	0	0	0	0	0	0	0	0,011542	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
д	0,008373	0	0	0	0	0,007779	0	0	0	0	0	0,001445	0	0	0	0,001404	0	0,001439	0	0	0	0	0	0	0,001461	0	0	0	0	0	0	0		
е	0	0	0,001452	0,004329	0,007036	0,00133	0	0,001358	0	0,006969	0,005584	0,007011	0,002831	0,009224	0	0,001379	0,00417	0,004187	0,008527	0	0	0,001418	0	0	0	0	0	0	0	0	0	0		
ж	0,001397	0	0	0	0,006364	0	0	0	0,001399	0	0	0	0	0,001349	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
з	0,004237	0,002927	0,001445	0	0	0	0,001397	0,00145	0,002755	0	0	0,00131	0	0,004185	0,002875	0	0	0	0	0	0,00137	0	0	0	0	0	0	0,001409	0,001416	0	0	0		
и	0	0	0,002925	0,001353	0	0,007072	0	0,008591	0	0,002817	0,009897	0,005469	0,009989	0,006939	0	0	0,001358	0,00137	0,001383	0	0	0,004173	0,001388	0,001397	0,001482	0	0	0	0	0,001427	0,004763			
й	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0,001383	0	0	0	0	0	0	0,001411	0	0	0	0	0	0	0		
к	0,009862	0	0	0	0	0	0	0	0,007024	0	0	0	0	0,001326	0,013016	0	0,002806	0	0,002652	0,001344	0	0	0	0	0	0	0	0	0	0	0	0		
л	0,004967	0	0	0	0	0,001439	0	0,011339	0	0	0	0	0	0,007174	0	0	0	0	0	0	0	0	0	0	0	0	0	0,008321	0	0,002739	0,001356	0		
м	0,002852	0	0	0	0	0,004184	0	0,007103	0	0	0	0	0	0,00137	0,002822	0	0	0	0	0,001395	0	0	0	0	0	0	0	0	0	0	0	0,001402		
н	0,00788	0	0	0	0	0,011216	0	0,023248	0	0	0	0	0,001468	0,013388	0	0	0	0	0	0	0	0	0	0	0	0	0	0,001455	0	0	0,001416	0,001436		
о	0	0,004132	0,004341	0,010255	0,001461	0,007155	0	0,005494	0	0,004294	0	0,005595	0,002781	0,001436	0	0	0,005749	0,014079	0,009938	0	0	0	0	0,001404	0	0	0	0	0	0	0	0,001376	0,001344	
п	0	0	0	0	0	0,001395	0	0	0	0	0	0	0	0,011152	0	0,018165	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
р	0,015343	0	0	0,001388	0	0,011129	0	0	0,008452	0	0	0	0	0	0,007098	0	0	0,00131	0	0	0	0	0	0	0	0	0	0,001431	0	0	0	0	0	
с	0,001337	0	0	0	0	0	0	0	0,001402	0	0,002773	0,009155	0	0,001416	0,001418	0,002859	0	0	0,020975	0	0	0,001356	0	0,00148	0	0	0	0	0	0,001388	0,007054	0		
т	0,005714	0	0,004115	0	0	0,006935	0	0,005478	0	0,002774	0,00148	0	0,004287	0,017907	0	0,004161	0	0	0	0	0	0	0	0	0	0	0	0	0,005517	0	0,001455	0		
у	0	0	0	0	0,001344	0	0	0	0	0	0	0	0,0027	0	0	0,002675	0	0	0,001326	0	0	0	0	0	0	0	0	0	0	0	0	0		
ф	0	0	0	0	0	0	0	0,001397	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
х	0	0	0,001356	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
ц	0,001388	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
ч	0,00148	0	0	0	0	0,004292	0	0	0,001461	0	0	0	0	0	0	0	0	0	0,002719	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
ш	0	0	0	0	0	0	0	0	0,002842	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0,001482	0	0	0	
щ	0	0	0	0	0	0,001402	0	0	0,002744	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
ы	0	0	0,001431	0	0	0	0	0	0	0	0	0,00145	0,001455	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
ь	0	0	0	0	0	0	0	0,002824	0	0	0,001358	0	0	0,001383	0	0	0	0,005535	0	0	0	0	0	0	0	0	0	0	0	0	0	0,001416	0	
э	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0,002773	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
ю	0	0,00131	0	0	0,001429	0	0	0	0	0	0	0	0	0	0	0	0	0	0,002741	0	0	0	0	0	0	0	0	0	0,002744	0	0	0	0	0
я	0	0	0	0,001455	0	0	0	0	0	0	0	0,001436	0,001349	0	0	0	0,001416	0,004184	0	0	0	0	0	0	0	0	0	0,001402	0	0	0	0,001356	0	

Лабораторная работа №1

Произвольная часть текста:

м_я_бы_по

Использованные буквы:

Порядок n-граммы:

5 символов

15 символов

20 символов

25 символов

30 символов

35 символов

40 символов

45 символов

50 символов

Введенный символ:

Символ по счету:

Номер эксперимента: 51

Поле ввода символов:

Продолжить

Другой

Неравенство для энтропии:

2,76330013614539 < H < 3,46965350016743

Двоичная таблица угаданных символов:

00000000000000000001000000000000 ^

10000000000000000000000000000000

00000000000100000000000000000000

00000000000000000000000000000001

00001000000000000000000000000000 v

Вероятности:

q[1] = 0,36
q[2] = 0,14
q[3] = 0,04
q[4] = 0,04
q[5] = 0,02
q[6] = 0,02
q[7] = 0,02
q[8] = 0,02
q[9] = 0
q[10] = 0,02
q[11] = 0,02
q[12] = 0,04
q[13] = 0
q[14] = 0
q[15] = 0
q[16] = 0
q[17] = 0,04
q[18] = 0,02
q[19] = 0
q[20] = 0,04
q[21] = 0,02
q[22] = 0,04
q[23] = 0
q[24] = 0,02
q[25] = 0,02
q[26] = 0
q[27] = 0
q[28] = 0
q[29] = 0,02
q[30] = 0
q[31] = 0
q[32] = 0,04

Строка состояния:

Лабораторная работа №1

[illegible]



H	R
H1 with whitespaces	0,12618
H1 without whitespaces	0,114166
H2 with whitespaces	0,44269569
H2 without whitespaces	0,368318
H10	0,448>R>0,308
H20	0,564>R>0,434
H30	0,499>R>0,361

Під час виконання даної лабораторної роботи я навчився вимірювати частоти символів та біграм у тексті, визначати ентропію. Також, я навчився визначати надлишковість мови.